L Number	Hits	Search Text	DB	Time stamp
1	196	(caption or subtitle) same (extract$4 or captur$4) same (stor$4 or record$4)	USPAT	2004/05/11 16:10
2	0	(caption or subtitle) same (extract$4 or captur$4) same (stor$4 or record$4) same (still adj2 image$1)	USPAT	2004/05/11 16:11
3	6	(caption or subtitle) same (extract$4 or captur$4 or slic$4) same (stor$4 or record$4) same (index or (still adj2 (image$1 or picture$1)))	USPAT	2004/05/11 16:15
4	17	(caption or subtitle) same (extract$4 or captur$4 or slic$4) same (stor$4 or record$4) same (index or (still adj2 (image$1 or picture$1)))	US-PGPUB; EPO; JPO; DERWENT; IBM TDB	2004/05/11 16:16

DERWENT-ACC-NO: 1999-535784

DERWENT-WEEK: 199945

TITLE: Automatic index signal generating procedure for video recorder - involves detecting existence of subject-converting word from subtitle data extracted from video signal in order to generate corresponding index signal

PATENT-ASSIGNEE: NIPPON TELEGRAPH & TELEPHONE CORP[NITE]

PRIORITY-DATA: 1998JP-0029379 (February 12, 1998)

PATENT-FAMILY:

PUB-NO	PUB-DATE	LANGUAGE	PAGES	MAIN-IPC
JP 11234611 A	August 27, 1999	N/A	006	H04N 005/78

APPLICATION-DATA:

PUB-NO	APPL-DESCRIPTOR	APPL-NO	APPL-DATE
JP 11234611A	N/A	1998JP-0029379	February 12, 1998

INT-CL (IPC): G06F017/30, G11B015/02 , H04N005/78 , H04N005/91

ABSTRACTED-PUB-NO: JP 11234611A

BASIC-ABSTRACT:

NOVELTY - An index signal is generated if a subject-converting word is detected to exist within a subtitle data extracted from a video signal. DETAILED DESCRIPTION - An INDEPENDENT CLAIM is also included for a recording medium.

USE - For video recorder.

ADVANTAGE - Ensures automatic insertion of index signals.
DESCRIPTION OF
DRAWING(S) - The figure shows the block diagram illustrating the manner by
which the automatic index signal generating procedure is executed.

CHOSEN-DRAWING: Dwg.2/4

TITLE-TERMS: AUTOMATIC INDEX SIGNAL GENERATE PROCEDURE VIDEO RECORD DETECT
EXIST SUBJECT CONVERT WORD SUBTITLE DATA EXTRACT VIDEO SIGNAL ORDER
GENERATE CORRESPOND INDEX SIGNAL

DERWENT-CLASS: T01 T03 W04

EPI-CODES: T01-J05B; T03-E05; W04-B; W04-B04B5; W04-F;

SECONDARY-ACC-NO:
Non-CPI Secondary Accession Numbers: N1999-398487

PAT-NO: JP410232884A

DOCUMENT-IDENTIFIER: JP 10232884 A

TITLE: METHOD AND DEVICE FOR PROCESSING VIDEO SOFTWARE

PUBN-DATE: September 2, 1998

INVENTOR-INFORMATION:
NAME
FUKUDA, TORU
TSUCHIYA, HARUKI

ASSIGNEE-INFORMATION:
NAME COUNTRY
KK MEDIA RINKU SYST N/A

APPL-NO: JP09304862

APPL-DATE: October 20, 1997

INT-CL (IPC): G06F017/30, G11B019/02 , G11B027/00 , H04N005/765 , H04N005/781

ABSTRACT:

PROBLEM TO BE SOLVED: To exactly grasp an entire video software by detecting a position to change the state of video software constitutive data and generating the summary of video software by extracting several representative images from the video software based on that information.

SOLUTION: A superimposed character discrimination and storage part 4 discriminates and segments superimposed characters (caption). The information (timing information) of position to let the superimposed characters appear is

sent to a cut discrimination and storage part 1. The superimposed characters
are preserved on a hard disk 6 (or RAM) together with the information of this
position. The cut discrimination and storage part 1 receives the information
of position to let the superimposed characters appear and extracts a still
picture from the prescribed position of correspondent cut (scene). The hard
disk 6 stores the superimposed characters, still pictures and cut images
clipped and designated by a viewer during the reproduction of representative
images. According to a program stored in a ROM 3 and the hard disk 6, a CPU 2
executes processing for generating the summary.

(19)日本国特許庁（ＪＰ）　(12)公開特許公報（A）　(11)特許出願公開番号

特開平10－232884

(43)公開日　平成10年(1998)９月２日

(51)Int.Cl.[6]		識別記号	FI		
G06F	17/30		G06F	15/401	320A
G11B	19/02	501	G11B	19/02	501D
	27/00			27/00	E
H04N	5/765		G06F	15/40	370D
	5/781		H04N	5/781	510F

審査請求　未請求　請求項の数16　ＦＤ　（全 15 頁）

(21)出願番号　特願平9－304862

(22)出願日　平成９年(1997)10月20日

(31)優先権主張番号　特願平8－334479

(32)優先日　平８(1996)11月29日

(33)優先権主張国　日本（ＪＰ）

特許法第65条の２第２項第４号の規定により図面第２図の一部は不掲載とする。

(71)出願人　591015854
株式会社メディア・リンク・システム
東京都中央区東日本橋２－２－10　東日本橋オリモビル

(72)発明者　福田　徹
東京都中央区東日本橋２－２－10　株式会社メディア・リンク・システム内

(72)発明者　槌屋　治紀
東京都中央区東日本橋２－２－10　株式会社メディア・リンク・システム内

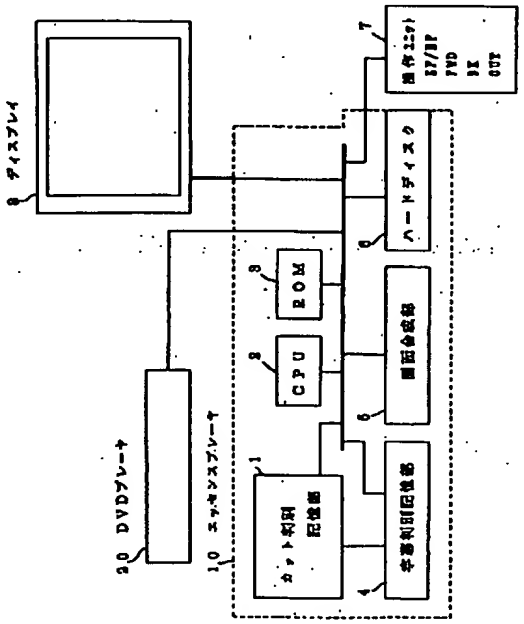
(74)代理人　弁理士　松井　晃一

(54)【発明の名称】　映像ソフトウェア処理方法及び映像ソフトウェア処理装置

(57)【要約】

【課題】　ドラマ等、映像ソフトウェアの全体像を例えば５分間で的確に把握可能とする。

【解決手段】　画像、音、字幕コードその他、映像ソフトウェア構成データが変化する位置に着目し、例えばＤＶＤに収録された２時間ドラマの内容を、当該各変化した位置付近を代表する各代表画像の集合（要約）で表現する。これを１画面づつ所望の間隔でコマ送り表示する。本発明では、音、文字、画像の変化や出現、画像の差分などを検出し、前記代表する画面抽出の動機とする。



1

【特許請求の範囲】

【請求項1】　画像、音、字幕その他の映像ソフトウェア構成データの少なくとも一つについて、その状態が変化する位置を検出し、該検出された位置の情報に基いて、前記映像ソフトウェアから幾つかの代表画像を抽出し、前記映像ソフトウェアの要約を生成することを特徴とする映像ソフトウェア処理方法。

【請求項2】　画像、音、字幕その他の映像ソフトウェア構成データの少なくとも一つについて、その状態が変化する位置を検出し、該検出された位置の情報を、幾つかの代表画像からなる前記映像ソフトウェアの要約を生成する為の位置情報として、前記映像ソフトウェアに付加することを特徴とする映像ソフトウェア処理方法。

【請求項3】　画像、音、字幕その他の映像ソフトウェア構成データの少なくとも一つについて、その状態が変化する位置であるとして前記映像ソフトウェアに付加されている位置の情報を読み出し、該位置情報に基いて、前記映像ソフトウェアから幾つかの代表画像を抽出し、該映像ソフトウェアの要約を生成することを特徴とする映像ソフトウェア処理方法。

【請求項4】　画像、音、字幕その他の映像ソフトウェア構成データの少なくとも一つについて、その状態が変化する位置を検出し、該検出された位置に係る幾つかの代表画像を抽出し、要約として前記映像ソフトウェアに付加することを特徴とする映像ソフトウェア処理方法。

【請求項5】　画像、音、字幕その他の映像ソフトウェア構成データの少なくとも一つについて、その状態が変化する位置を検出し、該位置に係る代表画像として抽出し、前記映像ソフトウェアに付加されている要約を読み出して、順次再生することを特徴とする映像ソフトウェア処理方法。

【請求項6】　前記生成された要約、或いは付加されていた要約の再生時、視聴者から命令があった場合、当該再生していた代表画像を抽出した位置付近から当該映像ソフトウェアを通常状態で再生することを特徴とする請求項1、請求項3又は請求項5の何れかに記載の映像ソフトウェア処理方法。

【請求項7】　通常状態での映像ソフトウェアの再生時、視聴者から命令があった場合、該位置付近から当該映像ソフトウェアの要約生成を実行することを特徴とする請求項1、請求項3、請求項5又は請求項6の何れかに記載の映像ソフトウェア処理方法。

【請求項8】　画像、音、字幕その他の映像ソフトウェア構成データの少なくとも一つについて、その状態が変化する位置を検出する位置検出手段と、該検出された位置の情報に基いて、前記映像ソフトウェアから幾つかの代表画像を抽出し、前記映像ソフトウェアの要約を生成する要約生成手段とを備えたことを特徴とする映像ソフトウェア処理装置。

【請求項9】　画像、音、字幕その他の映像ソフトウェア構成データの少なくとも一つについて、その状態が変化する位置を検出する位置検出手段と、該検出された位置の情報を、前記映像ソフトウェアの要約を生成する為の位置情報として、前記映像ソフトウェアに付加する位置情報付加手段とを備えたことを特徴とする映像ソフトウェア処理装置。

【請求項10】　画像、音、字幕その他の映像ソフトウェア構成データの少なくとも一つについて、その状態が変化する位置であるとして前記映像ソフトウェアに付加されている位置の情報を読み出す位置情報読み出し手段と、該位置情報に基いて前記映像ソフトウェアの要約を生成する要約生成手段とを備えたことを特徴とする映像ソフトウェア処理装置。

【請求項11】　画像、音、字幕その他の映像ソフトウェア構成データの少なくとも一つについて、その状態が変化する位置を検出する位置検出手段と、該検出された位置の情報に基いて、前記映像ソフトウェアから幾つかの代表画像を抽出し、該代表画像を要約として前記映像ソフトウェアに付加する画像付加手段とを備えたことを特徴とする映像ソフトウェア処理装置。

【請求項12】　画像、音、字幕その他の映像ソフトウェア構成データの少なくとも一つについて、その状態が変化する位置を検出し、該検出された位置の情報に基いて、前記映像ソフトウェアから抽出され、要約として前記映像ソフトウェアに付加されている代表画像を順次再生する再生手段を備えたことを特徴とする映像ソフトウェア処理装置。

【請求項13】　前記生成された要約、或いは付加されていた要約の再生時、視聴者から命令があった場合、当該再生していた代表画像を抽出した位置の付近から当該映像ソフトウェアを通常状態で再生する再生手段を備えたことを特徴とする請求項8、請求項10又は請求項12の何れかに記載の映像ソフトウェア処理装置。

【請求項14】　通常状態での映像ソフトウェア再生時、視聴者から命令があった場合、当該再生していた位置付近から前記要約生成を実行する要約手段を備えたことを特徴とする請求項8、請求項10、請求項12又は請求項13の何れかに記載の映像ソフトウェア処理装置。

【請求項15】　各カットを代表する夫々のカット代表画像のヒストグラムについて多次元空間内での距離を算出し、該距離が近いもの同士を纏めてグループを形成し、該各グループを代表する各位置を抽出することで、前記構成データの状態が変化する位置の検出を実行することを特徴とする請求項1乃至請求項7の何れか一に記載の映像ソフトウェア処理方法。

【請求項16】　各カットを代表する夫々の画像のヒストグラムについて多次元空間内での距離を算出する距離算出手段と、該距離が近いもの同士を纏めてグループ化するグループ化手段と、該各グループを代表する位置を

検出するグループ代表検出手段とで、前記位置検出手段を実現することを特徴とする請求項8乃至請求項14の何れか一に記載の映像ソフトウェア処理装置。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は映像ソフトウェア処理方法及び映像ソフトウェア処理装置に関し、詳しくは、例えばDVD（デジタルビデオディスク）に収録された2時間ドラマの内容から、的確な代表画像を抽出し、例えば1画面数秒づつ順に静止表示することで、当該映像ソフトウェアの全体を、例えば5分間等の短時間で、高速且つ的確に把握可能にする為の映像ソフトウェア処理方法、及び映像ソフトウェア処理装置に関する。

【0002】始めに用語の使い方について断っておく。先ず、本明細書に於て映像ソフトウェアとは、DVD収納映像、デジタルTV（テレビジョン）放送、地上波TV放送、インターネットやパソコンで利用される映像ファイル、コンピュータグラフィック、映画などを言い、DVD、ハードディスク、磁気テープ、フィルムその他、そのデータ形式や媒体の形式如何に拘らず、或る程度の数の画像（フレーム）の連続で、被写体の動きや、情景、その場の雰囲気、その他を視聴者（利用者）に訴える為のもの全てをいう（或る程度としたのは、4コマ漫画などを除く意。）。

【0003】また本明細書に於て、各請求項を始めとして何箇所かでは、例えば「位置の情報」と「位置情報」というように、間に「の」の字が有るか無いかの違いのみの言葉を何個か使用している。これは、前後の関係で語呂を良くし読みやすくする為に使い分けただけのことで、意味は同じである。判り切ったことかも知れないが念のため断っておく。「シーン」と「カット」、「視聴者」と「利用者」、「画像」と「静止画」、「画面」等についても同様とする。更に、用語は適宜略して使用することがある。例えば、映像ソフトウェアを「映像ソフト」と、本発明に係る処理を含む動作を「エッセンスプレイ」と、その処理により生成される要約物を「エッセンス」、「ビデオエッセンス」或いは「VE」というように言うことがある。

【0004】

【従来の技術】現代社会はマルチメデイア時代に向かっている。これらの時代を支える情報産業には下記のようなものがある。

情報1次産業＝作家、新聞記者、カメラマン等（情報生産）

情報2次産業＝出版・新聞社、レコード会社、放送局等（情報加工）

情報3次産業＝書店、新聞配達、レンタルビデオ等（情報流通）

情報の加工・流通では自動化が進展している。このため、情報生産に携わる人々が増大しつつある。

【0005】こうした情報産業を支えるのがエレクトロニクス産業で、情報産業の高度化に貢献するさまざまなハードやソフトが提供されつつある。この結果として、以下のような状況が生み出される。

・情報生産者が増え、玉石混交のコンテンツが大量に生産される。

・マルチチャンネルTV、DVD等、新メディアの開発による映像情報洪水。

・個人の情報摂取時間には限界があり、対応しきれなくなる。

【0006】またテレビジョンの変貌もある（楽しむから探すへ）。即ち、

・チャンネルが少ない時代は家族全員で共通の映像を楽しんでいた。今は、

・マルチチャンネル時代であり、自分が見たい映像を探すパーソナル化が進み、テレビ雑誌による番組選択、ザッピング（リモコンによるチャンネルの頻繁な切り替え）、マルチ画面の利用が盛んになって来ている。

【0007】ところで、映像は時間の流れに沿って見るもので、これを見るのには、それなりの時間をとる必要がある。しかし、上述したようなのような世の中の動きの中で、映像に対する時間節約の現状を見てみると、

・ビデオでタイムシフト・・・見たい時、（暇な時）にその録画を見る。

・ビデオで時間短縮・・・・・早送り、ダイジェストプレイで要点を見る。

・2画面ビデオ ・・・・・・・スポーツと音楽を同時に楽しむ。一方を再生しつつ他方を受信する。

などの対応が見られる（ビデオ：本来は「映像」の意であるが、ここでは、ビデオデッキや録画のこと、或いはこれらを使用することを指す）。

【0008】上述のとおり、映像は時間消費的である。しかし、人間の時間は限られている。もし映像の要点をどこででも簡単に短い時間で把握することが出来れば、多くの映像の中から自分の本当に見たい映像（作品、番組その他）のみを選択し、ゆっくりと自分のペースで見ることが出来る。このような願望は、マルチメディア時代と言われる以前からも存在した。従来は、その目的のため、上述の早送りなどが利用されてきた。また近年では、上記ダイジェストプレイと言われる機能を搭載したビデオデッキ、即ち、約1．5～2倍の早送り画面に、通常の再生速度の音を併せて再生し、通常の約半分の時間で内容を知ることが出来る機能を搭載したものも市販されている。

【0009】

【発明が解決しようとする課題】しかし、前記早送り、ダイジェストプレイ等、従来の手法は、何れも煩雑で視聴者（利用者）にとって極めて不満足なものであった。即ち、これら従来の手法は、例えばダイジェストプレイによれば音こそ通常の速度で再生可能であるものの、画

面は全て早送りであった。このため、映像はブレて見にくく、しかもそれが当該映像ソフトのポイントを表わすものかどうかを視聴者自身が判断しなければならないから、目を皿のようにして画面に注目しつつ（ダイジェストプレイの場合は耳の方にも）、早送り、再生、レビュー、一時停止、といった操作を頻繁に行なわなければならなかった。この手法では、特定のシーン（カット、場面）だけ見つけ出したいなら兎も角、映像ソフトの全体像を掴もうとするには、目も神経もかなり疲れる。

【００１０】また、前述のように、大量の映像情報が供給されるようになりつつある、或いは先端技術の時代になって来つつあると言われても、そのことによって人間の能力や自由な時間が大幅に増えて行くという訳ではない。この為、

・忙しいのでゆっくり映像ソフトを見ていられない。

・大量の映像ソフトの中から見たいものを素早く取り出したい。

・自分のペースで短い時間に自由に多くの映像ソフトを見たい。

という新たな要求も出てきている。

【００１１】本発明の目的は、このような人間の時間節約ニーズに答え、従来の手法と異なる新たな手法で、映像ソフトの全体の的確な把握を可能にする手法を提供することにある。

【００１２】

【課題を解決するための手段】上記目的達成のため本発明では、画像、音、字幕その他の映像ソフトウェア構成データの少なくとも一つについて、その状態が変化する位置を検出し、該検出された位置の情報に基いて、前記映像ソフトウェアから幾つかの代表画像を抽出し、前記映像ソフトウェアの要約を生成する（請求項１）。また、画像、音、字幕その他の映像ソフトウェア構成データの少なくとも一つについて、その状態が変化する位置を検出し、該検出された位置の情報を、幾つかの代表画像からなる前記映像ソフトウェアの要約を生成する為の位置情報として、前記映像ソフトウェアに付加する（請求項２）。

【００１３】また、画像、音、字幕その他の映像ソフトウェア構成データの少なくとも一つについて、その状態が変化する位置であるとして前記映像ソフトウェアに付加されている位置の情報を読み出し、該位置情報に基いて、前記映像ソフトウェアから幾つかの代表画像を抽出し、該映像ソフトウェアの要約を生成する（請求項３）。また、画像、音、字幕その他の映像ソフトウェア構成データの少なくとも一つについて、その状態が変化する位置を検出し、該検出された位置に係る幾つかの代表画像を抽出し、要約として前記映像ソフトウェアに付加する（請求項４）。

【００１４】また、画像、音、字幕その他の映像ソフトウェア構成データの少なくとも一つについて、その状態

が変化する位置を検出し、該位置に係る代表画像として抽出し、前記映像ソフトウェアに付加されている要約を読み出して、順次再生する（請求項５）。また、請求項１、請求項３又は請求項５の何れかに記載の映像ソフトウェア処理方法に於て、前記生成された要約、或いは付加されていた要約の再生時、視聴者から命令があった場合、当該再生していた代表画像を抽出した位置付近から当該映像ソフトウェアを通常状態で再生する（請求項６）。また、請求項１、請求項３、請求項５又は請求項６の何れかに記載の映像ソフトウェア処理方法に於て、通常状態での映像ソフトウェアの再生時、視聴者から命令があった場合、該位置付近から当該映像ソフトウェアの要約生成を実行する（請求項７）。。

【００１５】また、画像、音、字幕その他の映像ソフトウェア構成データの少なくとも一つについて、その状態が変化する位置を検出する位置検出手段と、該検出された位置の情報に基いて、前記映像ソフトウェアから幾つかの代表画像を抽出し、前記映像ソフトウェアの要約を生成する要約生成手段とを備える（請求項８）。また、画像、音、字幕その他の映像ソフトウェア構成データの少なくとも一つについて、その状態が変化する位置を検出する位置検出手段と、該検出された位置の情報を、前記映像ソフトウェアの要約を生成する為の位置情報として、前記映像ソフトウェアに付加する位置情報付加手段とを備える（請求項９）。

【００１６】画像、音、字幕その他の映像ソフトウェア構成データの少なくとも一つについて、その状態が変化する位置であるとして前記映像ソフトウェアに付加されている位置の情報を読み出す位置情報読み出し手段と、該位置情報に基いて前記映像ソフトウェアの要約を生成する要約生成手段とを備える（請求項１０）。また、画像、音、字幕その他の映像ソフトウェア構成データの少なくとも一つについて、その状態が変化する位置を検出する位置検出手段と、該検出された位置の情報に基いて、前記映像ソフトウェアから幾つかの代表画像を抽出し、該代表画像を要約として前記映像ソフトウェアに付加する画像付加手段とを備える（請求項１１）。

【００１７】また、画像、音、字幕その他の映像ソフトウェア構成データの少なくとも一つについて、その状態が変化する位置を検出し、該検出された位置の情報に基いて、前記映像ソフトウェアから抽出され、要約として前記映像ソフトウェアに付加されている代表画像を順次再生する再生手段を備える（請求項１２）。また、請求項８、請求項１０又は請求項１２の何れかに記載の映像ソフトウェア処理装置に於て、前記生成された要約、或いは付加されていた要約の再生時、視聴者から命令があった場合、当該再生していた代表画像を抽出した位置の付近から当該映像ソフトウェアを通常状態で再生する再生手段を備える（請求項１３）。また、請求項８、請求項１０、請求項１２又は請求項１３の何れかに記載の映

像ソフトウェア処理装置に於て、通常状態での映像ソフトウェア再生時、視聴者から命令があった場合、当該再生していた位置付近から前記要約生成を実行する要約手段を備える（請求項14）。

【0018】また、請求項1乃至請求項7の何れか一に記載の映像ソフトウェア処理方法に於て、各カットを代表する夫々のカット代表画像のヒストグラムについて多次元空間内での距離を算出し、該距離が近いもの同士を纏めてグループを形成し、該各グループを代表する各位置を抽出することで、前記構成データの状態が変化する位置の検出を実行する（請求項15）。

【0019】そして、請求項8乃至請求項14の何れか一に記載の映像ソフトウェア処理装置に於て、各カットを代表する夫々の画像のヒストグラムについて多次元空間内での距離を算出する距離算出手段と、該距離が近いもの同士を纏めてグループ化するグループ化手段と、該各グループを代表する位置を検出するグループ代表検出手段とで、前記位置検出手段を実現する（請求項16）。

【0020】（作用）即ち本願発明では、画像、音、字幕コードその他、当該映像ソフト構成データが変化をする位置に着目し、例えばDVDに収録された2時間ドラマの内容を、当該各変化位置付近を代表する各画像の集合（要約）で表現することとし、具体的には、これを1画面数秒づつ順次静止表示して（コマ送り表示して）、例えば5分間で当該ドラマ等の全体像を的確に把握出来るようにする。本発明では、音、文字、画像の特徴などをさまざまに利用し、またコンピュータによる画像処理技術を利用して、この要約をつくりだす。このシステムを利用すると、例えば図2の30分の長さの「ちびまるこちゃん」の映像ソフトVSは、その右に示される何枚かの画像VDと字幕CPからなる要約に纏められる（なお図2の絵は、漫画「ちびまるこちゃん」から引用。）。

【0021】

【発明の実施の形態】以下、本発明の詳細を説明する。＊理解を容易にするため、先ず図1に示す実施の形態の一例であるエッセンスプレーヤ10について説明し、その後、本発明の種々の展開について説明をする。即ち、エッセンスプレーヤ10は、デジタルビデオプレーヤ20に接続して使用され、カット判別記憶部1、字幕判別記憶部4、画面合成部5、ハードディスク6、CPU2、ROM3、及び操作ユニット7、ディスプレイ8等を備えている。

【0022】各部の機能は以下のとおりである。

字幕判別記憶部4：字幕（キャプション）を判別し、切り出す。字幕が出現する位置の情報（タイミング情報）をカット判別記憶部1に送る。字幕はこの位置の情報と共に、ハードディスク6（またはRAM）に保存される。

カット判別記憶部1：字幕の出現する位置の情報を受け、それに対応するカット（シーン）の所定位置から、静止画を取り出す（静止画＝一つの画面（1フレーム）。請求項にいう代表画像にあたる。）。静止画にはアドレスが付けられ、ハードディスク（またはRAM）に保存される。

【0023】また、所定位置とは、例えば図3のPCの如き位置（シーン中央）をいう。シーンが切り替えられて少し経って、そのシーンを代表する画面（画像）が出現することが多いと推定し、ここでは、シーン中央を所定位置に挙げたが、ジャンルによってその特徴的画像の出現位置は異なるから、この所定位置は任意に定める。視聴者或いは映像ソフト製作者が一度要約抽出を試行し、その結果を見て所定位置を決めても良い。

【0024】ハードディスク6：字幕、静止画及び代表画像再生中に視聴者が切り抜き指定したカット映像が記憶される。

CPU2、ROM3：ROM3及びハードディスク6に格納されたプログラムに従い、CPU2が本発明に係る処理を実行する。

【0025】操作ユニット7：下記各種ボタンを備えている。

EP/NP	エッセンスプレイとノーマルプレイの切り替え	
FWD	エッセンスプレイ中、	次画面送り
BK	〃	前画面戻り
CUT	〃	ハードディスク6への切り抜き保存

ディスプレイ8：映像ソフトまたは代表画面（要約）の表示。

【0026】静止画と字幕には、その元となった映像の位置情報、及びこれら相互の関係を示す情報とが付され、リンクが張られる（情報相互の関連づけが行なわれる）。位置情報は、例えばDVDの格納アドレス、先頭から数えた画像（フレーム）の番号、先頭からそこまでの通常再生時間などで表現される。これに基いて、例えば、上記 EP/NP ボタンによるエッセンスプレイとノーマルプレイの切り替え動作、及び当該静止画抽出位置付※近の1画面または所望長さの連続画面の切り抜き保存が実行される。

【0027】なお、エッセンスプレイとノーマルプレイの切り替え動作は、例えば、静止画＋字幕の再生（エッセンスプレイ）から、利用者が見たいカットが見つかったら、その命令に従い、そのカット位置からノーマルプレイ（通常の再生）に移る為の処理、或いは、ノーマルプレイ中に、利用者の命令に従って、その位置から静止画＋字幕の再生処理に移行するという形でも実行される。また、エッセンス利用者は、自分が字幕を読むスピ

ードに合わせ、画面送りの速度を調節して見ることができる。

【0028】エッセンスプレイ中にCUTボタンを押されたときは、それに対応する1画面或いは所望長さのシーン（カット）が映像データとしてハードディスク6に保存される。この画像データにはファイル名が付される。字幕からテキストが取り出せるときは、それをファイル名とするのも良い。そうすれば後で検索しやすい。また順にファイルを指定することにより、保存されたシーンを自由に繋いで編集可能にしておくのも良い。

【0029】字幕は文字コードで映像ソフトに添付されている場合と、映像そのものとして画像データの中に組み込まれている場合とがある。前者の場合は、当該媒体の格納フォーマットに従って、その文字コードを読み出せば良い。その出現位置（タイミングデータ）も当該格納位置に関連して容易に把握できる。字幕だけでもストーリーは理解出来るから、取り出した字幕データをそれぞれ1枚づつの静止画に仕立て、字幕だけの要約とするのも面白い。

【0030】映像そのものとして画像中に字幕が組み込まれている場合には、近年その精度が上がって来た漢字OCRの手法を用い、字幕の有無を検出する。画面中に何か文字がある、という程度の認識が出来れば十分であるから、例えば、一般に字幕表示位置とされる、画像の下部とか両端部分に着目し、その画像を粗い解像度で捉え、そこから何らかの文字が読取れるかどうかで、字幕の有無を判別すれば良い。

【0031】以下、映像の種類毎に、映像、カット、字幕の関係を例示する。

［1］従来の字幕付き映画

従来の字幕付き映画は、フィルムに字幕が焼き付けてある。従ってその処理方法は上述したようになり、C＝カット（シーン），S＝字幕，G＝静止画又は短い動画とすると、夫々の関係は、例えば下記のようになる。

映像＝C1＋C2＋C3＋・・・・・・・・・・Cx

カットは数秒～数十秒で、その中に字幕がついている場合と字幕なしの場合がある。字幕は1カットに1回の場合と数回の場合がある。

カット	字幕（スーパーインポーズ）
C1	S11＋S12
C2	0
C3	S31＋S32＋S33
・	・
・	・
・	・
Cx	Sx1＋Sx2＋Sx3＋Sx4

【0032】この場合、エッセンス画像の抽出は、例えば以下の如く行なう。

1.字幕を判別し取り出す。

2.字幕を表示順に列べる。

S11＋S12、S31＋S32＋S33、・・・Sx1＋Sx2＋Sx3＋Sx4

3.各カットで最初に字幕が表示された位置の静止画を1枚切り出す。

C1	G1
C2	－
C3	G3
・	・
・	・
・	・
Cx	Gx

4.静止画と字幕とに夫々アドレスを持たせ、ハードディスク6に保存する。

【0033】これらとエッセンスプレイとの関係は、以下の如くである。

1.DVDやデジタルビデオで通常再生映像を見ているときに、エッセンスプレイボタン（EP/NP ボタン）を押すことにより、静止画と字幕が表示される。

2.例えばC3のカットをプレイ中に、エッセンスプレイボタンを押すと、

G3＋S31

が表示され、次ページボタン（FWD ボタン）を押すことにより、

G3＋S32、G3＋S33・・・Gx＋Sx1、Gx＋Sx2・・・

とエッセンス画面が順次表示される。また戻りボタン（BK）を押すと1画面前に戻ることができる。

3.エッセンスプレイ中にノーマルプレイボタン（EP/NP ボタン）を押すと、そのカットの初めから通常の映像を見ることができる。

4.エッセンスプレイ中に切り抜きボタン（CUT）を押すと、そのカットの映像をハードディスク6に保存することができる。

【0034】［2］映像＋字幕（画像形式の字幕）

（1）映像と字幕の構成。

映像と字幕が別の画面として構成され、再生時に合成表示されるものもある。この場合の関係は下記のようになる。

映像＝C1＋C2＋C3＋・・・・・・・・・・Cx

字幕＝S1＋S2＋S3＋・・・・・・・・・・Sy

カット	字幕（スーパーインポーズ）
C1	S1＋S2
C2	0
C3	S3＋S4＋S5
・	・
・	・
・	・
Cx	Sy

【0035】（2）エッセンス画像の抽出方法

1.字幕を表示順に列べる。

S1、S2、S3、S4・・・Sy

2.各カットで最初に字幕が表示された静止画を１枚切り出す。

C1	G1
C2	－
C3	G3
・	・
・	－
・	・
Cx	Gx

3.静止画と字幕にそれぞれアドレスを持たせ記憶装置に保存する。

【００３６】（３）エッセンスプレイ

1.ＤＶＤやデジタルビデオで映像を見ているときに、エッセンスプレイボタンを押すことにより、静止画と字幕が表示される。

2.例えばC3のカットをプレイ中に、エッセンスプレイボタンを押すと、

G3＋S3

が表示され、次ページボタンを押すことにより

G3＋S4、G3＋S5・・・

とエッセンス画面を見ていくことができる。戻りボタンを押すと１画面前に戻ることができる。

3.エッセンスプレイ中にノーマルプレイボタンを押すと、そのカットの初めから通常の映像を見ることができる。

4.エッセンスプレイ中に切り抜きボタンを押すと、そのカットの映像を記憶装置に保存することができる。

【００３７】［３］映像＋字幕（テキストデータ）

字幕が画像データでなくテキストデータの場合、テキストデータを一旦画像に展開して字幕にし、再生時に映像と合成表示する。この場合も、字幕が表示されるカットが指定されているので、映像から字幕判別の必要がない。［２］のときと同様にエッセンスプレイ処理、操作が行える。

【００３８】以上、実施の形態例について説明をした。本発明は更に広範な形式で実施できる。以下、これらについて詳述する。なお項目番号は、ここから改めて付ける。先ず、ＶＥは原映像から抽出するものである。そこで原映像の特性の把握と、そこからどのような要約映像をつくりだせるかということが問題になる。先ずこの点から説明する。

【００３９】１．原映像（マルチメデイア・データ）の要素

ＶＥを抽出するための原映像は、ＤＶＤ、テレビ、ビデオ、パソコンなどに使われるマルチメデイア映像であり、大きく分けると次の２種類がある。

１）ストリーム映像

映画、ビデオなど、従来から存在する映像のこと。一定の流れに従って変化する。

２）リンク映像（ハイパージャンプ映像）

近年、ゲームなどで使用されている映像で、ユーザーの選択により、リンク先が変り、ストーリー展開が変化する映像。ＶＥはこれら２種類の何れも対象にするが、現在は、「ストリーム映像」の方が多数なので、当面、こちらが対象になる。

【００４０】これら映像ソフトには、以下のような要素が含まれている。

１）文字コード（ＴＣ）

文字コードは画面の説明として、難聴者のために必須となる。これから製作されるマルチメデイアデータの多くに含まれるようになる。

２）イメージ文字（ＴＩ）

文字として見えるが、文字コードになっていないもの。この種の文字は、自然の映像ではなく文字であるかどうかを知ることができれば利用できる。この処理は、前述の如く既存の文字認識手法で行なえる。

【００４１】３）静止画像（Ｓ）

画面に一定時間静止した映像として表示されるもの。ＴＶなどにおける動画映像の中で変化のないもの、またはパソコンなどに利用される静止画ファイルである。

４）動画映像（Ｄ）

映画、ビデオなどの映像そのもので、連続した多数のフレームで被写体の動きなどを表わすもの。ＴＶの場合には１秒間に３０枚の画像になる。画面の動きが早いと一枚の静止画では解像度が低下することがある。その場合は、静止画に替え、１秒ほどの動画映像を切り出す。これも請求項にいう「代表画像」に含むものとする。

【００４２】５）音（Ａ）

・音には、音声とそれ以外の音とがある。

・音声は人間の声であり、はっきりした意味のあるものである（広義には、「音声」には全ての「音」が含まれるが、ここでは、一応、前述の如き狭義で使用する（厳密な区別はしていない））。

・「それ以外の音」には、例えば、音楽、効果音、周囲の音などがある。音声とそれ以外の音とは、マルチメディアデータではあらかじめ分離されて格納されるのことが多い。

【００４３】２．ＶＥの構成

ＶＥの構成要素としては、上記マルチメデイア要素を全て利用可能である。しかし、短時間にこれを見るためには、静止画と文字とを使い、必要に応じて短い動画映像を使用するのが効果的である。

１）静止画・・・・ユーザーは、これをページめくりするようにして見る。

２）文字・・・・・文字コードデータがある場合にはこれを表示する。イメージ文字がある場合にはこれを画面に貼り付ける。画面に静止させるか、例えば左右に流れるように表示すると良い。

３）動画映像・・・ユーザがページ送りすると開始され

る。
4）音・・・・・・ユーザがページ送りすると開始される。ページ送りが早いときは出力しなくて良いであろう。
【0044】3．VEを見る場合のユーザーの操作
1）ユーザーはVEを本のページをめくるようにしてみる。
2）ひとつのページは静止画または動画である。
3）ページには音がついていて、ユーザーがそのページを開けば、前述のように音の出力が行われる（繰り返しもできる）。
4）自動めくりモードも用意しておくと良い。例えば、静止画のとき、視聴者が設定した間隔、例えば数秒ごとに自動的に次のページに進むようにすると良い。動画映像はページが開かれるとすぐに動き出し、終了すると次のページへ進むようにすると良い。再生速度を上げると音声は聞き取りにくいから、音の出力は効果音を中心にする方が良いであろう。
5）原映像へのリンク（ハイパージャンプ）
VEを見ていてその原映像を見たくなったら、ボタンを押す。これだけで原映像にリンクされ、もう一度押すとまたVEに戻ることもできる。
【0045】4．ユーザーが設定出来るパラメータ
VEを利用するときに、ユーザーが設定できるパラメータとして、例えば下記のようなものが考えられる。
1）要約率の大きさ
2）自動めくり／手動めくり
3）要約の手法（代表画像抽出の動機）
デフォルトの要約法を定めておくのも良い。それ以外にも視聴者が選択出来るようにしておくのが良い。複数の動機を使って、例えば人の声が存在する位置と、歓声の上がった位置の両方のついて代表画像を抽出しても構わない。
【0046】［動機の例］
・シーンの切り替わり
・一定時間ごとの周期的切り出し
＊・拍手、歓声などの音のクライマックスの切り出し
ビデオなどにVEボタンをつけると、ユーザーはこのボタンを押してVE映像をすぐに取り出せるようになる。パソコンではVEボタンをクリックすると、VE映像のばらばらめくりができる。
【0047】5．VEの抽出方法
VEの抽出には、マルチメデイアとしての情報の総合的な関係を利用する。静止画や動画映像を取り出すタイミングの発見に、動画映像の状態、音、文字などの情報を利用する。先に述べたことと重複するが、整頓の意味で改めて説明すると、VEを抽出する方法として、例えば次のような手法が考えられる。
【0048】1）先頭画面の切り出し
原映像の開始時の画面をタイトル（代表画面）として単純に切り出す（抽出する）。
2）周期的切り出し
一定時間ごとの周期的な画面の切り出し（要約率に応じて10秒おきなど）。
3）カットの切り出し
ひとつのシーンをとりだし、そこから1枚の静止画を切り出す。
4）字幕の切り出し
画面の一部に字幕があればこれを静止画として切り出す。
【0049】5）音声のあるシーンの切り出し
音声のないシーンを飛ばして、音声のあるシーンのみを切り出す。
6）文字パターン画面
文字情報が画面に明示されたシーン（パターン表示など）を1代表画面として切り出す。
7）クライマックス
拍手、大歓声など音声が大きくなったクライマックス時の短時間の動画映像を切り出す。
【0050】6．想定される対象とVEの抽出
各種ストリーム映像について、VEの具体的抽出動機（要素の変化位置）を例示すれば以下のようになる。

映像のジャンル	VE抽出動機
ニュース	パターン（フリップ）のあるシーン
ドラマ	字幕のあるシーン
	音声のあるシーン
ドキュメンタリー	音声のあるシーン
英会話	字幕のあるシーン
スポーツ	拍手、歓声の上がるシーンとその周辺（音声クライマックス）
アニメ	字幕のあるシーン
	長く静止しているシーン
	音声のあるシーン
TVショッピング	字幕のあるシーン（価格などの情報が見える）
歌番組	音楽の始まるシーン

（音声から判別）

教育番組	パターンのあるシーン
バラエテイショー	歓声の上がるシーン
オーケストラ	音楽のスタートシーン

（周期的抽出）

天気予報	静止したシーン

（パターン（フリップ）＝文字や画を書いた板のことで、ＴＶで話し手などが使用する。）

【００５１】７．要約率

要約率＝ＶＥを標準の速度の自動めくりで見る時間／原映像を見るのに要する時間

である。ＶＥを利用すると、原映像の表示時間の１／１０～１／３０の時間（要約率）でその内容を見ることが出来る。

【００５２】ＶＥを見るのに必要な時間はユーザーの操作に依存する。ユーザーはゆっくりとページをめくるように見てもいいし、素早く映像を送って見てもよいし、また自動ページめくりで見てもよい。自動めくりの場合で、２時間（１２０分）の原映像から抽出されるＶＥは、おおよそ４～１２分程度になる。

【００５３】８．ＶＥの実現形態

ＶＥ技術の実現形態は、例えば、以下の各種が考えられる。

１）映像ソフトへの付加

ＶＥが標準化された場合には、どのような映像にも、一般的にＶＥが標準的に付加されることになる。映像ソフト制作側は、制作段階でＶＥのデータを付加するようにする。

【００５４】２）ハードへの組み込み

ＶＥは、様々なハード（装置）に組み込むことも出来る。専用ＬＳＩチップとして生産し、販売しても良い。

３）ＶＥをソフトとして利用する

ＶＥ技術をコンピュータソフトウェアとして実現し、パソコンなどに組み込んで実行する。利用者はオプションにより、さまざまな映像に対してＶＥを使いわけることができる。インターネットからの映像の取り出し時などにも利用できる。

【００５５】本発明は、更にマルチ・メディア以外の既存の映像ソフトに対しても適用できる。以下、それらについて説明する。先ず、最終的に取り出される要約映像は以下のように構成される。

（１）静止画像

（２）短い動画映像

（３）文字（映画の字幕にあたる）

（４）音

【００５６】（１）原映像の入力装置

通常の映像取り込装置、例えばビデオデッキ、レーザディスク用デッキなどを利用する。

（２）要約処理の方法

１）簡便法

＊・原画像がＮ枚の画像で構成され、これからＭ枚の画像を抽出して要約を生成するとするならば、Ｎ／Ｍ枚ごとに映像を静止画としてとりだす。これは最も単純な方法である。

・音は、無音部分を取り除き、音声が言葉になっている部分をとりだす。言葉になっている部分と音楽、騒音（轟音）などは周波数分析により比較的判別可能である。取り出される静止画の映像の付近にある音声のみをとりだす。

・文字データが与えられている場合には、静止画付近に表示される文字を取り出して、静止画に重ね合わせる。

【００５７】２）差分法（シーン切り替えタイミングの発見）

これは、時間的に大きな変化をする映像を見つけ出し（シーンの切り替えなど）、これに注目して映像を取り出すものである。

・シーンが切り替わったら、次にシーンが切り替わるまでの時間を調べ、その中間の時刻の画面を静止画としてとりだす（図３ＰＣの位置）。あるいは、そのシーン切り替えが生じる中間の部分のある時間幅に相当する動画像を要約画像としてとりだす。

・文字の扱い

マルチメディア・データの場合には、取り出した静止画像／動画像の付近にある文字を「文字データ」から取り出して、静止画にスーパーインポーズする。そうでないときは字幕を判別して表示する。

・音声の扱い

取り出した静止画像／動画像の付近にある音声を「音声データ」から取り出して、静止画が映っているあいだその音声を一回だけ、またはくりかえし出力する。

【００５８】画像の処理方法としては以下のような手法を利用できる。

１）ピクセルの集約処理

映像は２次元的な広がりとして例えば３００ｘ２６０というような点の集まりである。それぞれの点に色がついている。テレビの映像ではこのような映像が１秒間に３０枚必要である。この映像のピクセル構成は、この後の処理に対して通常は多すぎるので、ピクセルの集約化を行う。例えば４ｘ４の点で集約すれば、原映像は１／１６のデータ量にすることができる。８ｘ８で取り出せば、１／６４に集約できる。

【００５９】このようにすると、

① ズームアップ／ダウンのとき
② ゆっくりと変化する画像
に対しても、小さい変化を切り捨てる効果を持つ。このような前処理を行った上で　夫々の値、
a（t,x,y）
をとりだす。但し、t：時刻　x，y：集約した映像の座標　a：その点（x，y）の色の値であり、a＝R＋G＋Bなどとするのが一般的である（R，G，Bは3原色情報の値）。

【0060】2）時間方向の集約化処理
時間的に変化する映像データは、そのまま扱うには冗長であり、またムダも多い。そこで、前記前処理を行なった各点のデータa（t,x,y）に対して、以下の何れかの処理をする。
（1）周期的なサンプリング
（2）映像の時間的な差分比較処理
一枚の画像のデータを構成する上記各点のデータ、a（t，x，y）について、時間的な差分を求める。
d（t，x，y）＝a（t，x，y）－a（t－△t，x，y）
但し、△t：適宜の時間幅

【0061】画像全体（x，y）についてこの値を集計する。式で表わせば、
【数1】
$$Da(t)=\sum_{x,y}(d(t,x,y))$$
となる。これは、時間的に△tだけ隣り合う2枚の画像の間の変化量（差分）を示している。

【0062】ここで値が大きい順に、Da（t）をN個とりだす。このようにして取り出したN個のDa（t）の例を図3に示す。値が大きい位置、即ち映像の差分が大きい位置であるCS1，CS2は、そこで画像に大きな変化があること、即ちそこでシーンが切り替わっている可能性が高いことを表わしている。そこで、このCS1とCS2の間を一つのシーンと考え、この中から1枚の代表画像を抽出する。ジャンルによって異なるが、シーンを端的に表わす静止画は、一般にシーン中央付近にある。そこで、この図3の例では、シーンの中間位置PCを当該代表画面を取り出す位置としている。

【0063】なお映像ソフトを構成する各フレームについて、例えば夫々の中央付近の水平走査線1本分の画像データに着目して画像と画像の間の変化の度合い（シーンの切り替わり）を検出するようにしても良い。具体的には、例えば、この水平走査線1本分の画像データを、N個の区間に分け、夫々の区間について平均値を求める。そして、各区間毎に、その前の画像の当該区間の平均値との差分を求める。この差分を各フレーム毎に総和し、その値が大きくなっている位置、即ち、図3のCS1、或いはCS2に当たる位置を求め、これを上記同様のシーンの切り替え位置であるとする。

【0064】要約映像は次のような特徴をもつ。
① ユーザーが好きな時間のタイミングでページをめくるようにして見ることができる。
② 静止画の場合には単純である。
③ 動画の場合には、ある部分がひとつのページ内で短い時間に動画として表示され、指示により繰り返して見ることができる。

【0065】以下、変形例について説明する。先ず、要約再生の際、それを構成する代表画面を1度に4つ、ディスプレイに表示しても良い（代表画面4つを合成し、1画面にして表示する）。代表画像は、もともと1つの映像ソフトから出発している。従って夫々は相互に関連を持っており（ストーリーを持っており）、これらが複数個、一度に画面に表示されても、視聴者は容易にその内容を感得出来る。従って、例えばN枚の代表画像を一度に画面表示したとすれば、単純には、1枚づつ静止表示のときのN分の一の時間で、当該要約を再生することが出来る。また、映像ソフトにストーリーがあることに鑑みれば、寧ろ一度に複数代表画像を表示する形式の方が、一つ一つ再生する形式より内容を感得しやすいかも知れない。

【0066】本発明は専用の装置において実現してもよいし、アプリケーションプログラムの一つとしてハードディスク等に格納しておき、必要に応じて呼出してコンピュータ上でも実施しても良い。フロッピーデイスクなどに格納して配布することも出来る。専用チップを製作してパソコンに組み込んだり、DVDプレーヤ、ゲームプレーヤに組み込んだり、VTRに組み込んだりするのも良い。パソコン等で実施する場合、画面上、或いはキーボードに、実施の形態例の「EP/NP」ボタン等を割り付けると良い。

【0067】一つの映像ソフトに複数の要約を付加しておき、視聴者が選択視聴出来るようにしても良い。例えば、字幕の出現位置を動機とした要約と、歓声の上がる位置を動機とする要約の二つを、当該映像ソフトに付加しておくのも良い。この処理は、位置情報を映像ソフトに付加する実施の形態に対しても適用出来る。また、位置の情報と要約画像の双方を映像ソフトに付加しておくのも良い。こうすると位置情報から当該要約を生成する機能が付けられていない通常の映像ソフト再生装置（DVDプレーヤ、ビデオデッキなど）でも、要約再生が出来る。

【0068】代表画面（要約）又は位置の情報が付加された映像ソフトは、DVD、ビデオテープの如き独立した媒体で伝達出来るほか、電波、ネットワーク、個別通信線その他の媒体を介しても伝達することが出来る。この際、当該代表画面抽出位置付近の種々のデータ（字幕、音等）も併せて伝達し、視聴者の選択で再生可能にしても良い。視聴者や管理者が選択した動機に基いて、映像ソフトから代表画像を抽出し、これらを、ハードディスクやビデオテープ等、適宜の記憶媒体に集積してお

くのも良い。こうすると、映像ライブラリー等に於て、目標とする映像ソフトの検索が、的確、且つ容易に実施できる。個人が所有する映像ソフトの管理にも活用できる。

【００６９】他の実施の形態例３０を図４～図１０に示す。この実施の形態例３０は、請求項１５の映像ソフトウェア処理方法及び請求項１６の映像ソフトウェア処理装置についての一つの実施の形態例であり、主としてＣＰＵ２，ＲＯＭ３上で実現される。

【００７０】例えばドラマに於て、或る場所（背景）に於て演技が行なわれ、カットが変って他の一つ或いは幾つかの場所でそれに続く演技が行なわれ、また元の場所に戻って演技が続けられるということは多い。要約は、それを構成する代表画像の数が適切であり、しかもそれらが多面的に夫々異なる場面を表現している方が当該映像ソフトの内容を感得しやすい。

【００７１】図４～図１０に示したこの実施の形態例３０は、このような要望に応え得るもので、各カットから取り出したカット代表画像について夫々そのヒストグラムを作成し、これらヒストグラムの多次元空間内での距離の遠近という視点から、各ヒストグラム、即ち各カット代表画像の近似性を求め、近似するもの、即ち前記距離が近いものを同じグループに纏めていく処理を繰り返して、これら多数のカット代表画像を、適宜数、例えば数十のグループに集約し、この各グループから代表画像を取り出すという処理を実行する。

【００７２】これにより、各代表画像の内容は、当該映像ソフトの異なった夫々の場面を表す多面的なものとなり、また、その数も適宜数に絞られる。また、ここで実施される各画面についての類似性判断は、映像ソフトの画面が一般的に備えている性質、即ち、カットが変るとそのヒストグラムの分布（波形、包絡線）がそこで目に見えて変化する、という性質を利用したものなので、その映像ソフトのジャンルが何であるかに拘らず、画一的、機械的に実施できるという利点がある。

【００７３】以下、この図４～図１０に示す実施の形態例３０について説明をする。見出しの番号はこの実施の形態例の説明で改めて付ける。

１）カットの切り出し。

映像ソフトは、多数のカットから成り立っている。一つのカットは数秒から数十秒の長さである。カットが変化する境界では、各フレームのデータが大きく変化する。＊これを検出する方法としては、先に説明した実施の形態例での判別法も一つであるが、ここでは、これに有効である各フレーム（画像）のヒストグラム（映像ヒストグラム）（Ｈ）を利用する手法を用いる。

【００７４】ヒストグラムの例を図５に示す。このヒストグラムはヒストグラム作成部３１で作成する。横軸に各画像の各ピクセルの色（Ｒ，Ｇ，Ｂ）の情報、或いは輝度や色合い（Ｙ，Ｕ，Ｖ）をプロットし、縦軸に一つの画像中での、その色等のピクセルの出現頻度（そのピクセル数）をプロットする。図５の例では、横軸に輝度Ｙを２５６段階でとり、縦軸に、その輝度Ｙを持つピクセルの発生頻度（ピクセル数）を示している。処理の簡素化のためには、横軸を１０～２０段階に区分して、図６に示すように纏めて表現するのが便利である。

【００７５】ヒストグラムは映像の持つ特性を簡潔に表現出来るので、様々に有効利用出来る。ここでは先ず前述のとおりカットの判別に利用する。即ち、差分算出部３２により、各ヒストグラムＨ（ｔ）の差分Ｈａを求め、その変化が大きいところを検出する。この差分が大きい位置がカットの切替り位置を表わす。

【００７６】式で表わすと、

Ｈ（ｔ）＝Ｈ（ｔ、ｙ１，ｙ２，ｙ３，…，ｙＮ）

但し、Ｈ（ｔ）：時刻ｔのフレームのヒストグラム

ｙｎ　：第ｎ区分の色または色や明るさ情報の頻度

（ｎ＝１、２、…、Ｎ。Ｎ＝横軸の区分数。）

【００７７】$\Delta Ha = H(t+1) - H(t) = \Sigma\ y_n(t+1) - y_n(t)$

（構成要素の差の絶対値の和）となる。図に表わすと、例えば図７のようになる。ここから一定値以上の変化のあるところを取り出せばカットの切り替わりを検出できる。

【００７８】２）カット代表ヒストグラムの抽出。

連続する映像、例えば１時間のドラマの連続した各画像を幾つものカットに切り出したので、次に代表ヒストグラム抽出部３３により、各カットの代表的なヒストグラム（Ｈｓ：代表ヒストグラム）を作る。これには、例えばカット内の初めの時刻のフレーム、中央付近の時刻のフレーム、終り時刻のフレームを取り出してヒストグラムを作成する。１カットの全体のヒストグラムについての時間平均を作成しこれを代表にしても良い。

【００７９】これを式で表わすと、

Ｈｓ　＝Ｈｓ（ｙｓ１，ｙｓ２，ｙｓ３，…，ｙｓＮ）

但し、Ｈｓ　：カットｓの代表ヒストグラム（ｓ＝１，２，…，Ｓ）

ｙｓｎ：第ｎ区分の色又は色や明るさ情報の頻度（ｎ＝１，２，…，Ｎ）となる。（そのカットの代表画面やカットの時間に亘る平均）

【００８０】３）カット代表ヒストグラムの近似度算出。

※次に、代表ヒストグラム間距離算出部３４で、各カット代表ヒストグラムＨｓ間の距離Ｄｓｔを算出する（ｓ，ｔ＝１，２，３，…，Ｓ）。距離Ｄｓｔは、或るカットｓの代表ヒストグラムと、別のカットｔの代表ヒストグラムとの間の距離であり、この値（Ｄｓｔ）は、夫々のヒストグラムの対応する各区分ｙｓｎ同士間の差の絶対値又は二乗値の合計として求められる。式で表わすと、

【数2】

$$Dst = \sum_{n=1}^{N} (ysn - ytn)^2$$

である。カット代表ヒストグラムの数がｓ個の場合、このようにして求められるそれら各代表ヒストグラム間の距離Ｄｓｔの数は、リーグ戦での試合数と同じで、ｓ×（ｓ－１）÷２個である。

【００８１】４）グループ化。

つぎに、カット間距離Ｄｓｔの小さなもの同士をとりまとめて、適当な数（Ｍ）のグループにする。この処理は、グループ化処理部３６で行なう。この処理には各種数学的手法が利用できるが、最も単純なのは以下のような方法である。先ずＤｓｔのうちで最小の距離にあるカット代表ヒストグラムｓとｔとを求める。そしてこのｓとｔを一つのグループにする。次に、残りの距離Ｄｓｔの中から最小の距離を持つｓとｔの組合わせを見つけてグループにする。

【００８２】当初は全カットの数に対応したｓ個のヒス＊

＊トグラムが独立に存在し、グループもこれと同じ数のｓ個存在する。ここで、グループ化の操作を１回行なうと、グループが一つ減少する。従って、このグループ化操作をＳ－Ｍ回繰り返せば、グループの数はＭ個になる。これにより、各カットの代表ヒストグラムｓ個を図８に示すような適切な数（Ｍ）のグループに纏めることが出来る。

【００８３】５）各グループの代表画像抽出。

グループの区分けが完了したら各グループから代表画像を抽出する。この処理は代表画像抽出部３７で実行する。この場合、夫々のグループから一つランダムにグループ代表画像を選んでも構わないし、厳密にするなら、グループの重心Ｈｇ（図８のｘ記号）を求め、この重心Ｈｇに最も近いものを代表画像とする。

【００８４】重心Ｈｇを求めるには、先ずそのグループに属することになった各ヒストグラムについて夫々のｙｓｎの値を平均する。式で表わせば、であり、

【数3】

$$ygn = \sum_{i=1}^{J} ysn(j) / J$$

但し、Ｊはグループ内のカットの数（ｊ＝１、２、…Ｊ）。

【数4】

$$Hg = Hg(yg1、yg2、\cdots、ygn、\cdots ygN)$$

と表わされる。

【００８５】そして各々のヒストグラムｊとグループ重心Ｈｇとの間の距離を求める。この距離が最小になるヒストグラム、即ちもっともグループ重心に近いヒストグラムの画像を、そのグループの代表画像として取り出す。これら代表画像は、その儘集合させ要約にしてもよい。ここまでの流れを纏めれば図９のようになる。

【００８６】この実施の形態例ではグループ化処理を行なった。従って、その処理結果を代表画像の抽出に反映させると、一層的確に原映像ソフトの内容を表現する要約が生成出来る。具体的には要約に含まれる夫々の画像に優先度を付け、この優先度に応じて下記の如く代表画像の選択を行なう。

【００８７】１）グループに含まれるカット（代表ヒストグラム）の数の反映。

グループを構成するカットの数が多い、即ち一本の映像ソフトの中で類似の画像なりカットなりが何度も出て来る場合、そのグループはその映像ソフトの中で存在意義が大きいと考えられる。そこで、このようなものについては、代表画像を一つだけ取り出すのではなく、頻度に比例した数の代表画像を取り出すことにするのが効果的と考えられる。この場合、取り出されたものが近似して※いては代表画像を複数にした意味が薄れる感じがする。従ってこのときは、なるべく距離の遠いものを選んで複数取り出すと良い。

【００８８】２）カットの合計時間の反映。

グループを構成しているヒストグラムの母体である各カットの持続時間の合計が長い場合も、そのグループも映像ソフトの中で存在意義が大きいと考えられる。そこでこれらの優先度を高くするために、グループ毎のカットの合計時間数に比例して幾つかの代表画像を取り出すのも効果的である。

３）頻度の低いカットの無視。

グループを構成するカットの数が少なければそのグループを無視し、その代表画像を要約に採用しないことも考えられる。

【００８９】代表画像の表示についても二つほど工夫がある。先ず、これらは大きな画面上に分散して配置して同時に多数表示するようにしても良い。言わば「鳥獣戯画」の如き表現形式である。こうすると、要約を時間的にではなく、空間的に把握することが出来るようになる。そして、この空間的に配置された代表画像の一つをクリックすると、その代表画像が動画として動きだすようにするのも良く、そこで更にダブルクリックすれば、

そのカットから原映像ソフトが再生されるようにするのも良い。

【0090】このとき、上記1），2），3）のような観点から、各代表画像の表示の大きさに重み付けを行なうと判りやすい。即ち図10に示すように、そのグループに属するヒストグラムが多いもの、或いはそのカットの持続時間の合計が多いものは表示面積を大きくし、且つ中央に配置する。こうすると、視聴者はその映像ソフトの言わばさわりを一瞥で感得出来る。なお、最初の実施の形態例についての変形例は、ここに説明した実施例にも当てはまる。

【0091】

【発明の効果】以上説明したように、本発明によれば、視聴者が欲する映像ソフト構成データに着目し、映像の要約を自動的に抽出することができる。視聴者は、それを自由に自分のペースで短時間のうちに見ることができ、当該映像ソフトの全体像を、簡単、かつ正確に把握することが出来る。これにより、簡単に大量の映像にアクセスできるようになる。

【0092】また、要約映像は、原映像に比較すればかなり小さいデータ量になり、通信や携帯端末にも適するものと成り得る。将来、光ファイバー高速通信網、大容量メモリー、高速CPUのおかげでデータ量そのものは問題にならなくなるとも思われるが、情報量自体が小さいことは、同じ通信手段でより沢山の種類の情報を伝達することが出来るという利点がある。また高速通信網への移行期でも、容易に実施出来るという利点もある。

【図面の簡単な説明】

【図1】実施の形態例を示すブロック図。

【図2】原映像ソフトウェアと抽出した要約（代表画面）の例を示す説明図。

【図3】各画像データ間の差分の例を示すグラフ。

【図4】他の実施の形態例を示すブロック図。

【図5】1枚の画像のヒストグラムの例を示すグラフ。

【図6】簡素化したヒストグラムの例を示すグラフ。

【図7】ヒストグラムの差分の例を示すグラフ。

【図8】n次元空間での各カット（ヒストグラム）のグループ化の例を示す概念図。

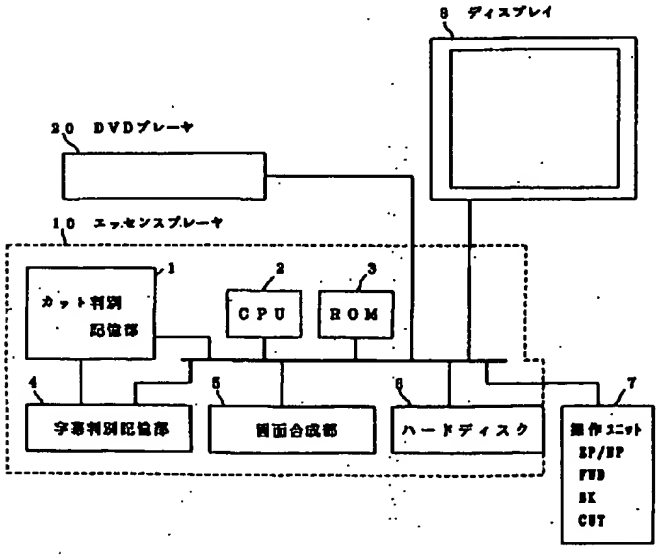
【図9】他の実施の形態例に於ける要約生成のプロセスを示す概念図。

【図10】空間配置の要約表示例を示す平面図。

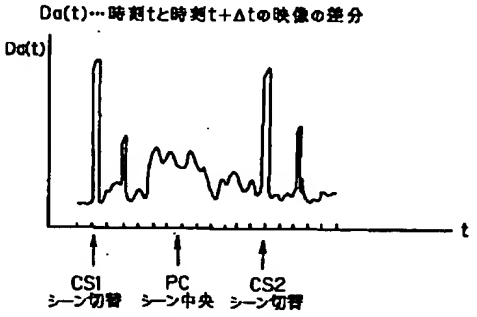
【符号の説明】

1…カット判別記憶部	2…CPU
3…ROM	4…字幕判別記憶部
5…画面合成部	6…ハードディスク
7…操作ユニット	8…ディスプレイ
10…エッセンスプレーヤ	20…DVDプレーヤ
30…他の実施の形態例	31…ヒストグラム作成部
32…差分算出部	33…代表ヒストグラム抽出部
34…代表ヒストグラム間距離算出部	36…グループ化処理部
37…代表画像抽出部	

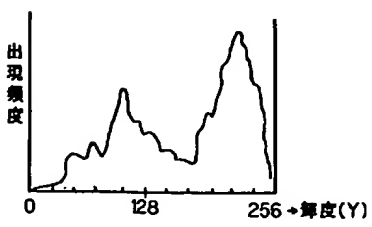
【図1】



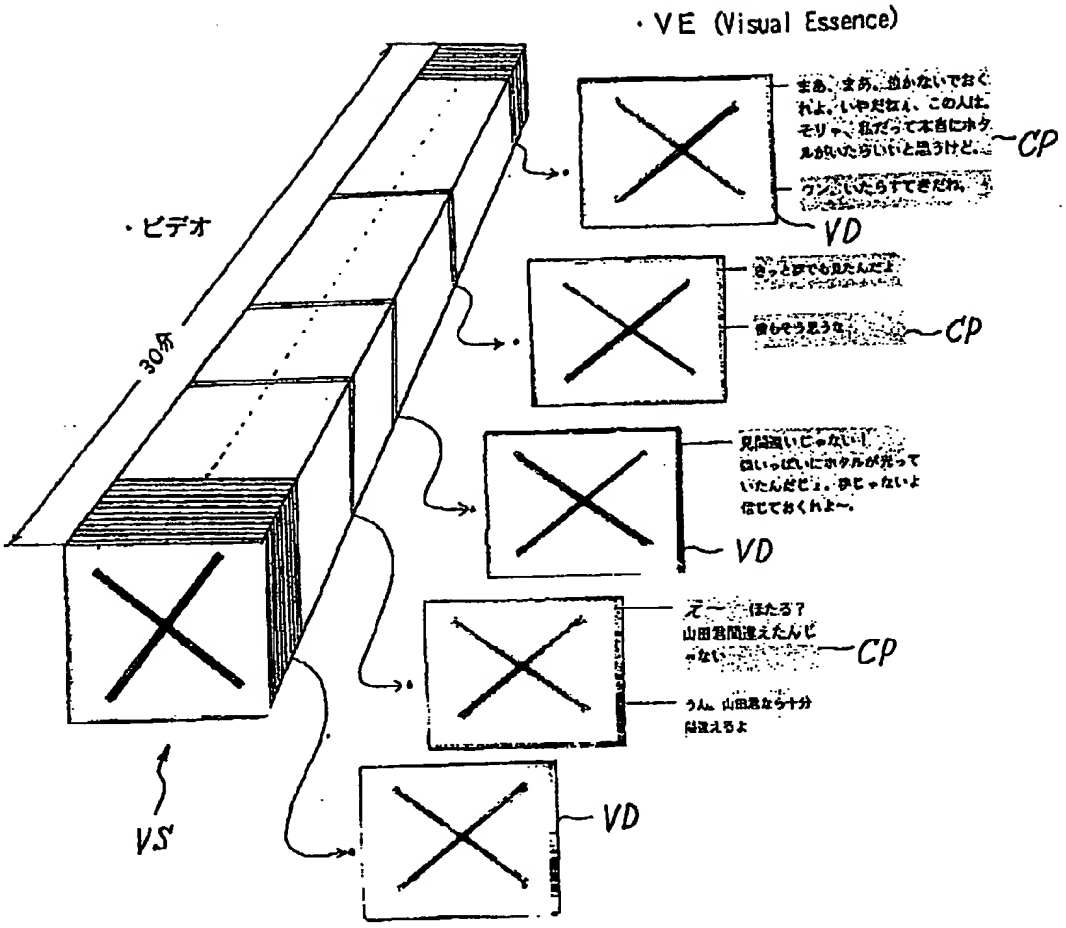
【図3】



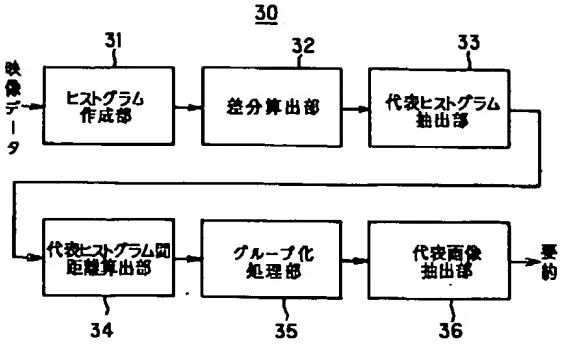
【図5】



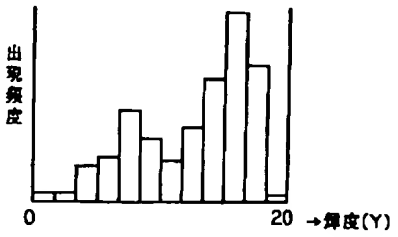
【図2】


・VE (Visual Essence)
・ビデオ
30分
VS
VD
CP
VD
CP
VD
CP
VD

【図4】


30
31
32
33
映像データ
ヒストグラム作成部
差分算出部
代表ヒストグラム抽出部
代表ヒストグラム間距離算出部
グループ化処理部
代表画像抽出部
要約
34
35
36

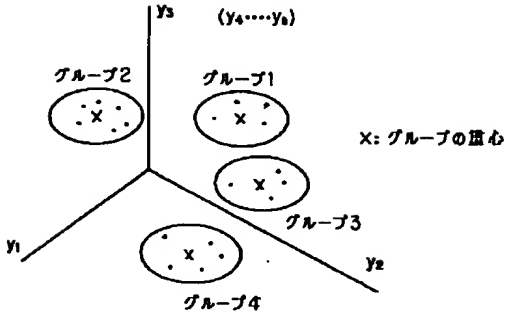
【図6】


出現頻度
0
20
→輝度(Y)

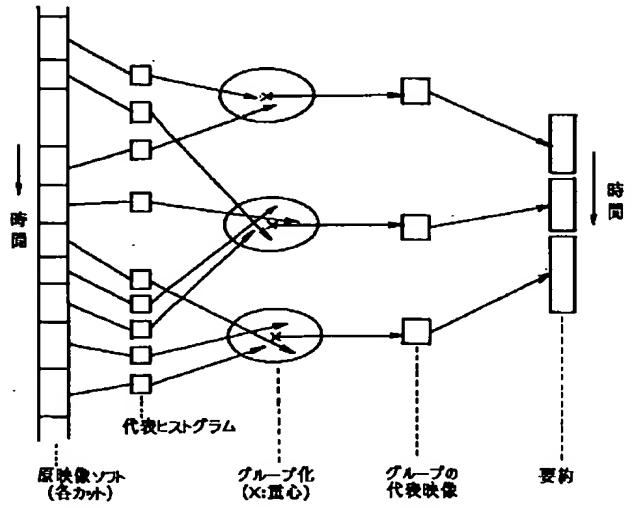
【図7】


ΔHα
t(時刻)
ひとつのカット
ひとつのカット

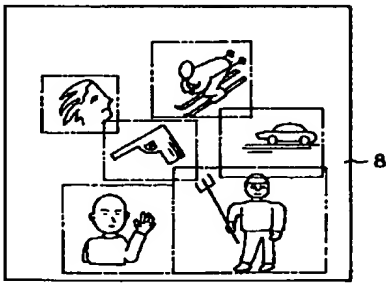
【図8】


y3
(y4……yn)
グループ2
グループ1
X: グループの重心
グループ3
y1
y2
グループ4

【図9】


時間
時間
代表ヒストグラム
原映像ソフト
(各カット)
グループ化
(X:重心)
グループの
代表映像
要約

【図10】


8